Attributes Related to Mortality Rate in the US (1960)

MATH 240- Final Project Paper

Instructor: Jane Sharman

Ruturaj Shete

Arkan Odeh

Neha Verma

Contents

Introduction

Mortality rate is defined as number of death in particular population based on the size of that population, per unit of time (It is typically expressed in units of deaths per 1000 individuals per year). Since, Mortality is considered to be one of the key ratio to evaluate any nations well-being and plays a dominant role in determining the growth of that nation. This paper examines the factors that affect mortality rate and how the mortality has changed over the last years.

When we relocated to San Francisco to pursue our graduate studies, we were given a task of selecting a data set for our statistical analysis. We came across with the data set related to the mortality rate in 60 metropolitan cities in the United States during the 1960; we realized that it will be interesting if we can compare the change in the mortality rate over the last fifty years. We also became curious to know about the factor that affects the mortality rate and how the change in environmental conditions has affected the mortality ratio before the 1960 and now.

The Mortality rate has been dependent on many factors in general but the most common factors includes medical facilities, health care, nutrition level, living standard, access to clean drinking water, hygiene levels and social factors such as conflicts and the level of violent crime. However, the data set we are using for our analysis consist of a combination of socio economic and pollution factors. These factors are very different to the ones mentioned above. Also over the last fifty years there has been lot of development in healthcare, bio-technology and overall improvement in standard of living, though the quality of the environment has decreased. So it will be interesting to see the change in the mortality rate due to the improvement in the factors affecting the mortality rate.

Objectives

The main objective of the paper is to determine the significant factor that affects the mortality rate during the 1960. It is very unlikely that only one factor would have influenced the mortality rate in the States. Therefore, it is important to analyze all the potential factors to determine the correlation. By the end of this analysis we will be able to compare the mortality rate before the 1960 and now.

Data Collection Methodology

This data is extracted from a research paper- "Instabilities of regression estimate relating air pollution to mortality", by McDonald G. C and Schwing R.C (1973), Technimetrics, Vol. 15.

The data which have been used to do the analysis have been collected from 60 metropolitan cities in United States in 1960. The original dataset can be seen in Appendix I. The data consist of one dependent variable (Mortality rate, expressed as death per 100,000 population) and 15 independents variables. The independent variables are very diverse as they are a combination of socio-economic factors and air pollution.

	Data Attributes
PREC	Average annual precipitation in inches
JANT	Average January temp. in degrees F
JULT	Average July temp. in degrees F
OVR65	% of 1960 SMSA population aged 65 or older
POPN	Average household size
EDUC	Median school years completed by those over 22
HOUS	% of housing units which are sound & with all facilities
DENS	Population per sq. mile in urbanized areas, 1960
NONW	% non-white population in urbanized area, 1960
WWDRK	% employed in white collar occupations
POOR	% of families with income < $3000
HC	Relative hydrocarbon pollution potential
NOX	Relative nitric oxides pollution potential
SO2	Relative Sulphur dioxide pollution potential
HUMID	Annual average % relative humidity at 1pm
MORT	**Total age-adjusted mortality rate per 100,000**

Table 1: Data Attributes

After compiling all the required data, we began by running correlation analysis to determine which variables were related to each other. Then we will do a massive regression on all the remaining significant independent variables, using mortality rate as our dependent variable. Then, we will do F test for the P-value in order to determine which variables adds to the model explanatory power or not, we will further eliminate the insignificant variables at 0.05 significant levels. After determining the significant variables, we will do the residual analysis to check for validity of the regression assumptions. Also after conducting the regression analysis and forming the regression equation we further plan to do a descriptive analysis of the significant variables to determine general measures of central tendency.

Assumptions

Since our main objective is to determine the significant factors which affect the mortality rate, we expected that there will be negative correlation between mean January temperatures in degrees of Fahrenheit and mortality rate. The base of our assumption is that human body has a tendency to maintain an inner core temperature of 98.7 degree Fahrenheit, and as temperatures decreases during the colder months, the body starts shutting down due to hyperthermia and human body start experiencing difficulty in breathing. Also people are more exposed to pollution during colder months, thereby increasing the mortality rate.

The second independent variables we expect to have a significant relation are Education (the median school years completed by those over 22 years). We assume mortality rate will have negative correlation with the education, as the educated population will have better income, they will have better affordability, which will affect their rate of survival and thus there will be a decrease in mortality rate.

Third factor that we assume will have a strong positive correlation with the mortality rate is the relative pollution potential of Sulfur Dioxide (SO2). It is believed that the emission of greenhouse gases due to human activities will change the earth’s climate, by causing changes in temperature, precipitation levels and weather variability.

The fourth variable which we think it will be significant is the population density. We assume as the population per square mile in the urbanized area will increase, there will be an increase in the level of pollution and hence the mortality rate will also increase accordingly.

Correlation Analysis

In order to check the correlation between our dependent variable and the independent variable we started with our first analysis i.e. the correlation analysis, in which we first did a VIF analysis to check the correlation between Mortality rate and one of the independent variables. Since, we know that if the value of VIF is greater than 5,

$$\text{VIF} > 5$$

Those candidates of independent variables are not significant. We found that the NOX (relative nitric oxides pollution) variable had a VIF value of 104.9. We reran the VIF analysis by eliminating the NOX independent variable under consideration, then POOR (% of families with income <$3000) variable which had a VIF value of 8.7 and also OVR65 (% of 1960 SMSA population aged over 65 years or older) with VIF value of 6.06. Therefore, we eliminated those 3 variables from our independent variables and went ahead and did Best Sub-set analysis, also called Cp value to see for the significant variables. We found that all the remaining 12 attributes had a Cp value which was less than or equal to the K+1value (where 'k' is the number of independent variables).

$$K+1 => Cp$$

Multiple Regression Analysis

After the correlation analysis we did the Multiple Regression analysis to see which all significant independent variables we have based on their p-value. So we eliminated 8 variables (JULT, POPN, PREC, HOUS, DENS, WWDRK, HC, HUMID), whose p-value was greater than 0.05 level of significance. We again re ran the regression and found these four significant independent variables at 0.05 level of significance; which are: SO2, JANT, EDUC, NONW. The dataset with these four significant independent variables can be seen in Appendix II.

Regression Statistics	
Multiple R	0.82270
R Square	0.67684
Adjusted R Square	0.65334
Standard Error	36.62578
Observations	60.00000

Table 2: Summary of multiple regression analysis

Coefficient of Multiple Determination-R^2

R-squared is the Coefficient of Determination, R^2 = 67.68%- means 67.68% of variation in the mortality rate can be explained by the variation in the independent variables (ln (Relative Sulphur dioxide pollution potential), Median school years completed by those over 22, Average January temperature and % non-white population in urbanized area, 1960).

ANOVA

	Df	*SS*	*MS*	*F*	*Significance F*
Regression	4	154528.01	38632.002	28.7987	6.331E-13
Residual	55	73779.64	1341.448		
Total	59	228307.64			

Table 3: Analysis of variance of multiple regression analysis

Significance F

The above table shows, that the p-value of the F-statistics is 633123-13 which is almost equal to zero and is less than 0.05 level of significance. Therefore, we know that there is a significant relationship between Mortality rate and one of the independent variable.

Now, when we looked closely at these four independent variables we found that individually all the 4 independent variables were significant based on their p-value.

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1135.730	71.1314	15.9667	2.63E-22	993.1799	1278.2809
ln(SO2)	8.971	3.5414	2.5333	1.42E-02	1.8743	16.0684
JANT	-1.414	0.5831	-2.4244	1.86E-02	-2.5822	-0.2451
EDUC	-21.149	6.0353	-3.5042	9.18E-04	-33.2437	-9.0539
NONW	4.723	0.6395	7.3855	8.81E-10	3.4412	6.0042

Figure 4: Multiple regression analysis

The p-value for the coefficient of January temperature is 0.018, which is less than 0.05 so there is a significant relationship between mortality rate and January temperature. The same way Education, Nonwhite population and Relative level of Sulphur dioxide do have a significant relationship with the mortality rate at 95% level of significance.

Confidence Level

The above multiple regression run under 95% confidence level. It means that 95% of all possible samples can be expected to be included in the true population parameter.

Range

The range for each variable is as follows:

1) Ln(SO2): 1.8743 to 16.0684
2) JANT: -2.5822 to -0.2451
3) EDUC: -33.2437 to -9.0539
4) NONW: 3.4412 to 6.0042

All the significant independent variables have a valid range, because they do not contain zero value. As all values are finite positive or negative numbers.

Final Multiple Regression Equation

Mortality rate = 1135.73 + 8.97 *ln (So2) - 21.15*(Education)-1.413*(January Temperature) + 4.72* (Non White population)

Where, Mortality rate is deaths per 100,000 populations

- Average January temperature is in degrees Fahrenheit
- Nonwhite population in urbanized area is in percentage
- Education is median school years completed by those over 22 population
- SO2 is relative Sulphur dioxide pollution potential

- Where, if we increase the median school years completed by those over 22 years of populations by one, we expect the mortality rate to decrease by 21.15 per 100,000 populations, on an average, keeping other attributes constant.
- Where, if we increase the relative Sulphur dioxide pollution by one percent, we expect mortality rate to increase by 0.0897 per 100,000 populations, on an average, by keeping other attributes constant.
- Where, if we increase the percentage non-white population in urbanized area by one percent, we expect mortality rate increase by 4.72 per 100,000 populations, on an average, by keeping other attributes constant.
- Where, if we increase the average January temperature by 1 degrees Fahrenheit, we expect mortality rate to decrease by 1.413 per 100,000 populations, on an average, by keeping other attributes constant.
- Y-coefficient is not significant to interpret, because it is not practical to make all the independent attributes zero. So y-coefficient has no practical interpretation.

Residual Analysis

In order to check for the assumptions of regression model, we went ahead and did residual analysis. The residual plots of the four significant variables can be seen below-

Residual Plots:

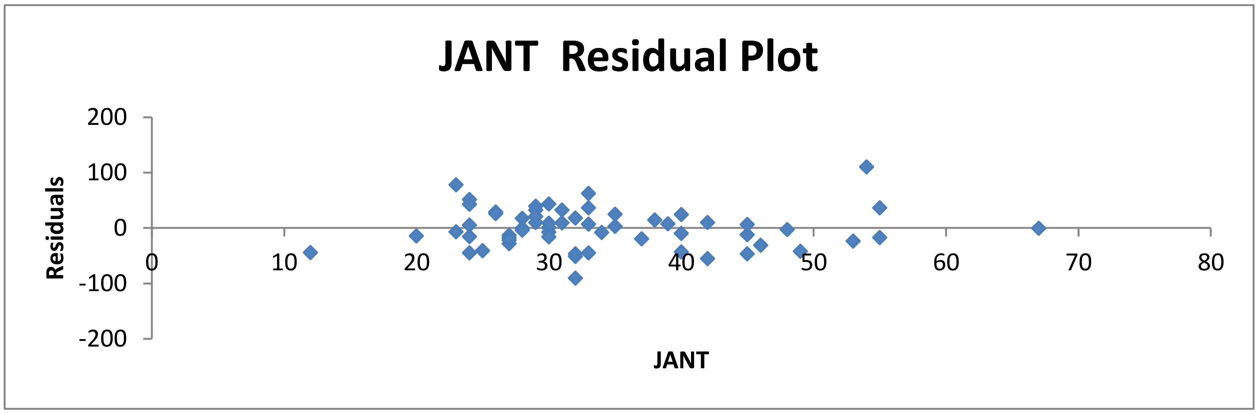


Figure 1: January Temperature Vs Residual Plots

The JANT vs. Residuals plot can be seen above. It looks random, there do not seem to be any violations of the regression assumptions. These obeys the following assumptions:

1) Linearity
2) Independence
3) Normality
4) Equal variance

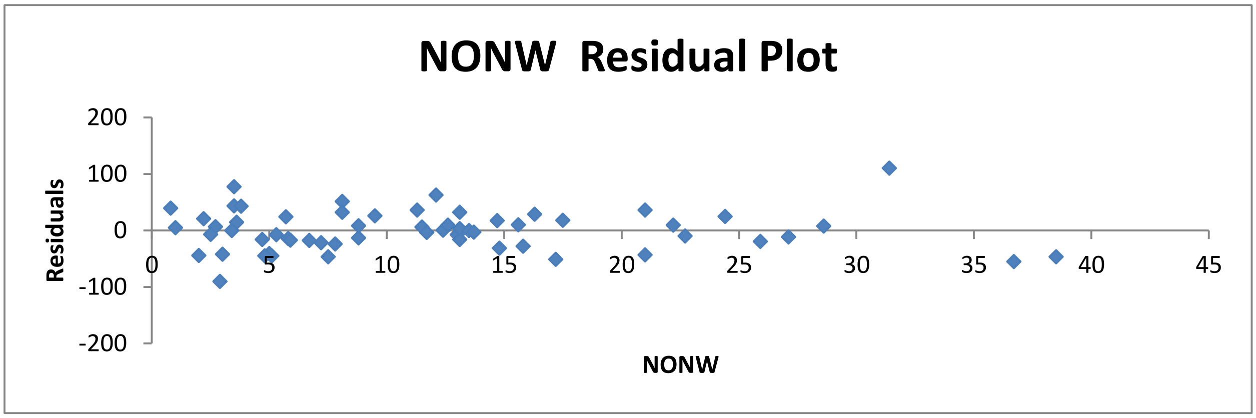


Figure 2: Nonwhite population Vs Residual Plot

The NONW vs. Residuals plot can be seen above. It looks random, and there do not seem to be any violations of the regression assumptions.

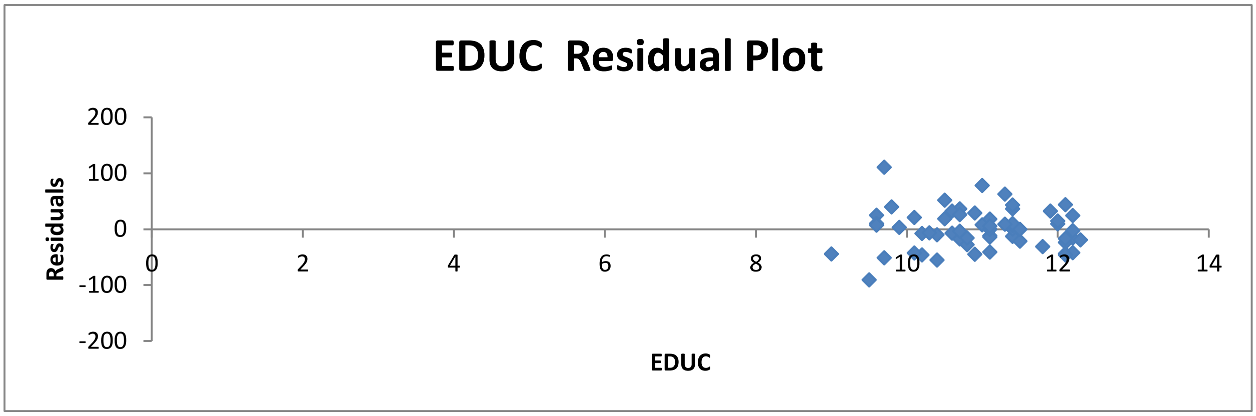


Figure 3: Education Vs Residual Plot

The EDUC vs. Residuals plot can be seen above. It looks random, and there do not seem to be any violations of the regression assumptions.

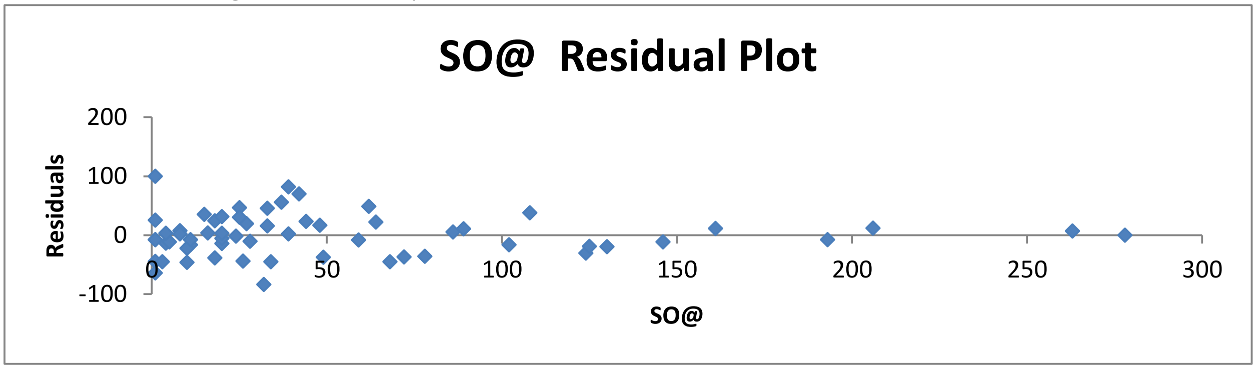


Figure 4: Sulphur dioxides Vs Residual plot

The SO2 vs. Residuals plot can be seen above. There seem to be a homoscedasticity issue with this plot and had violated the regression assumptions. Therefore, to solve this issue we used a natural log function, which is mentioned below:

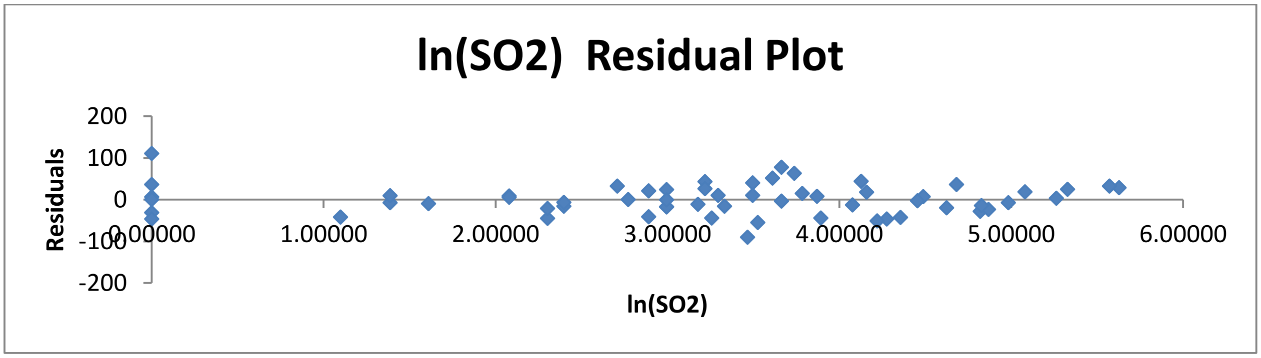


Figure 5: Natural logged Sulphur Dioxide Vs Residual plot

Now after using natural log function, the ln(SO2) vs. Residuals plot. It looks normal, and there do not seem to be any violations of the regression assumptions.

Descriptive Analysis

After we did the regression analysis and found the significant attributes we did the descriptive analysis to highlight the social economic factors that affect the mortality rate in 60 metropolitan cities during 1960 in the United States. The total quantitative analysis and five number summary is shown in Appendix III. The table below shows the highlighted attributes related to the social economic factors and pollution:

	Over 65	DENC	EDUC	HOUS	POOR	NONW	SO2	NOX	HC
Minimum	5.6	1441	9	66.8	9.4	0.8	1	1	1
Median	9	3567	11.05	81.15	13.2	10.4	30	9	14.5
Maximum	11.8	9699	12.3	90.7	26.4	38.5	278	319	648
Average	8.80	3876.05	10.97	80.91	14.37	11.87	53.77	22.65	37.85

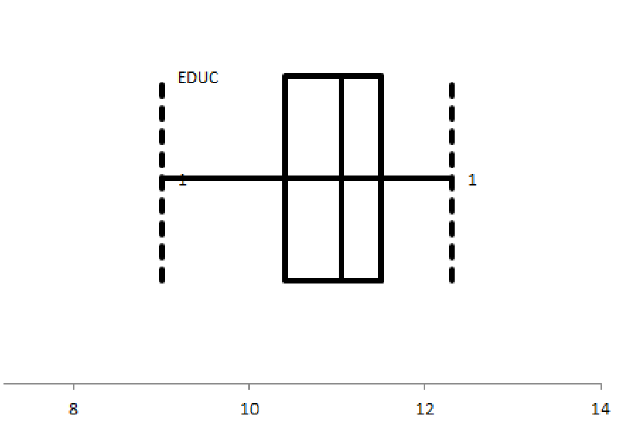
Table 5: Descriptive analysis of independent variables

As the analysis indicates Atlantic City had recorded the highest density of 9699 populations per square mile comparing the average density in all the 60 metropolitan cities of 3876 populations per square mile. The average of Non-white population was 11.89% and the average of population aged over 65 years was 8.8% from the entire population of the 60 cities.

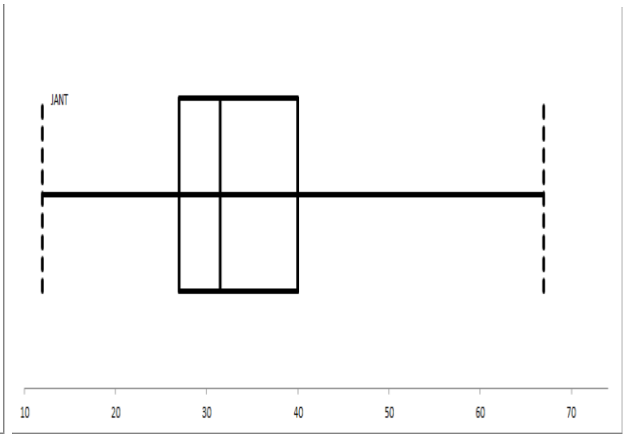
The average of the median school years completed by population aged over 22 years was 10.97 years but it is more left skewed which means the average population was more towards the lower numbers of years on the median school year completed. Although the average Percentage of poor in the cities was 14.37 % which represent family’s income below 3000$ per annum, the average percentage of the housing units were sound and with full facilities was 80.91%. Which was a good percentage as the minimum was 66.8% recorded in Fresno and the highest percentage was 90.7% recorded in Lansing.

As the regression analysis shows, the pollution indicators were significant as we had expected at the beginning of this paper but do not make a huge impact. This can be explained by the development in the industrial technology today than it was in 1960, which is mainly to meet the huge demand in the market. However, the average of the SO2 level of potentials pollution is 53.77 in the 60 cities with minimum level of 1 and maximum level of 278 which been recorded in Denver.

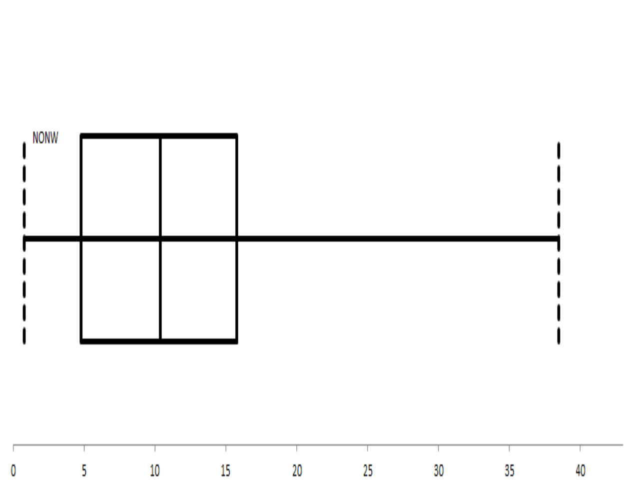
. EDUC: MEAN= 10.97



JANT: MEAN= 33.98



NONW: MEAN= 11.87



Ln(SO2): MEAN= 3.2

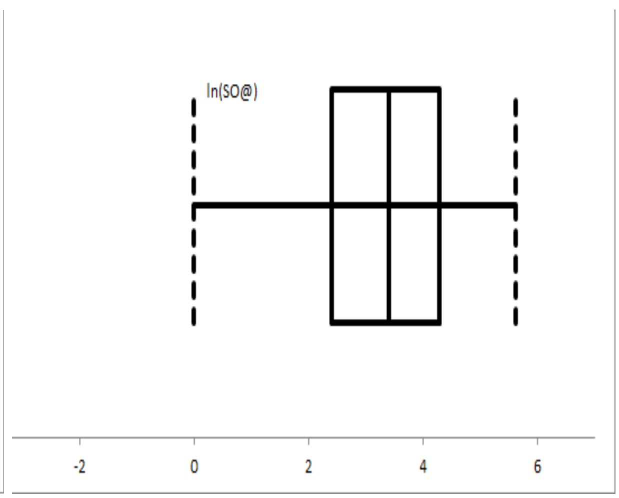
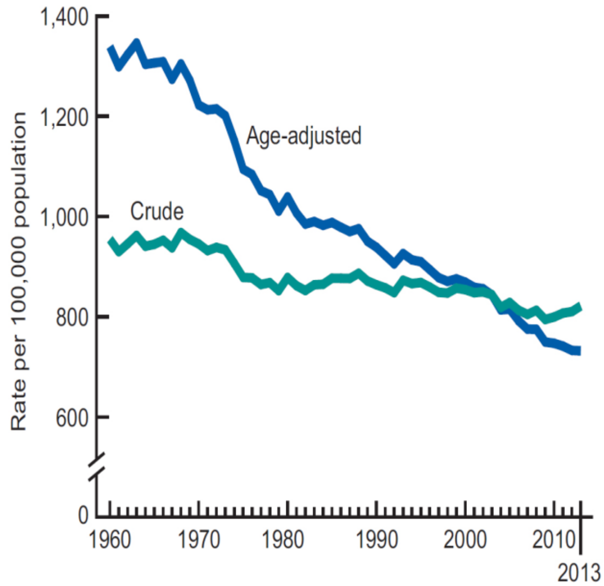


Figure 6: BOX plots of significant independent variables

Conclusion



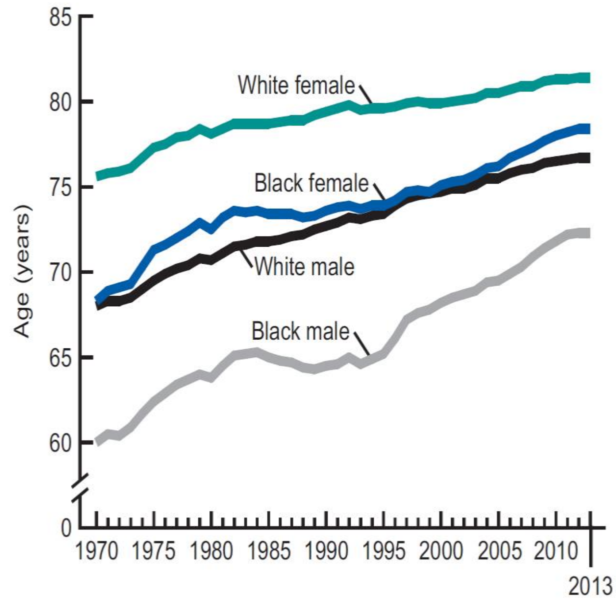


Figure 7: Mortality Rate 1960-2013   Figure 8: Life Expectancy 1970-2013

As our main objective in this paper is to identify the significant factors that had affected the mortality rate in the 1960 in the United States by collecting the data from 60 metropolitan cities across the country. From our regression analysis, we got four significant factors (education, Sulfur Dioxide (SO2), nonwhite and January temperature) which affected the mortality rate in 1960. This means that we were correct in our assumptions that there will be a positive correlation for SO2 and a negative correlation for both education and January temperature as significant factors to the mortality rate. However, we were surprised when nonwhite factor was shown significant rather than the population density as assumed.

This finding along with the descriptive analysis we had ran, highlighted some of the social economic factors that had affected the mortality rate in 1960 in the United State. The average education numbers of year completed in the middle school for the population for age under 22 years was 10.97 years, which indicate that most of the population on those cities did not finish their high school. This affected their annual income and the poorness level to an average of 14.37 with an annual income of less than $ 3000. As also can be seen from the analysis, nonwhite people were less privileged one during that time in terms of education, income, work environment and health support. This was the reason for increased mortality rate in 1960, as the nonwhite population in the cities was high during that time.

The mortality rate average in the 1960 was 940 deaths for every 100,000 people, whereas the new mortality rate in 2014 was 821 deaths for every 100,000 people according to the Department of Public Health. We had expected to find that the mortality rate will be higher in 2014 mainly due to the air pollution from the huge industrial development, excess use of automobiles, global warming affects and the chemical industry. However, the mortality rate has decreased over the last years and the average age has increased, this can be contributed to the technological development in health sector, education, more investments in clean energy and better income.

REFRENCES

McDonal, G.C. and Schwing, R.C. (1973). Instabilities of regression estimates relating air pollution to mortality, Technometrics, vol. 15, 463-482.

Deaths by age and age-specific death rates. (2000-2014). State of California, Department of Public Health, Death Records.

National Vital Statistics Reports. (2016, February 16). Deaths: Final data for 2013, Volume 64, Number 2.

Appendix I: Original Dataset

Cities	PREC	JANT	JULT	OVR65	POPN	EDUC	HOUS	DENS	NONW	WWDRK	POOR	HC	NOX	SO@	HUMID	MORT
Birmingham	36	27	71	8.1	3.34	11.4	81.5	3243	8.8	42.6	11.7	21	15	59	59	921.87
Mobile	35	23	72	11.1	3.14	11	78.8	4281	3.5	50.7	14.4	8	10	39	57	997.875
Montgomery	44	29	74	10.4	3.21	9.8	81.6	4260	0.8	39.4	12.4	6	6	33	54	962.354
Phoenix	47	45	79	6.5	3.41	11.1	77.5	3125	27.1	50.2	20.6	18	8	24	56	982.291
Little Rock	43	35	77	7.6	3.44	9.6	84.6	6441	24.4	43.7	14.3	43	38	206	55	1071.289
Fresno	53	45	80	7.7	3.45	10.2	66.8	3325	38.5	43.1	25.5	30	32	72	54	1030.38
Los Angeles	43	30	74	10.9	3.23	12.1	83.9	4679	3.5	49.2	11.3	21	32	62	56	934.7
Sacramento	45	30	73	9.3	3.29	10.6	86	2140	5.3	40.4	10.5	6	4	4	56	899.529
San Diego	36	24	70	9	3.31	10.5	83.2	6582	8.1	42.5	12.6	18	12	37	61	1001.902
San Francisco	36	27	72	9.5	3.36	10.7	79.3	4213	6.7	41	13.2	12	7	20	59	912.347
San Jose	52	42	79	7.7	3.39	9.6	69.2	2302	22.2	41.3	24.2	18	8	27	56	1017.613
Denver	33	26	76	8.6	3.2	10.9	83.4	6122	16.3	44.9	10.7	88	63	278	58	1024.885
Bridgeport	40	34	77	9.2	3.21	10.2	77	4101	13	45.7	15.1	26	26	146	57	970.467
Hartford	35	28	71	8.8	3.29	11.1	86.3	3042	14.7	44.6	11.4	31	21	64	60	985.95
New Haven	37	31	75	8	3.26	11.9	78.4	4259	13.1	49.6	13.9	23	9	15	58	958.839
Wilmington	35	46	85	7.1	3.22	11.8	79.9	1441	14.8	51.2	16.1	1	1	1	54	860.101
Washington	36	30	75	7.5	3.35	11.4	81.9	4029	12.4	44	12	6	4	16	58	936.234
Jacksonville	15	30	73	8.2	3.15	12.2	84.2	4824	4.7	53.1	12.7	17	8	28	38	871.766
Miami	31	27	74	7.2	3.44	10.8	87	4834	15.8	43.5	13.6	52	35	124	59	959.221
Orlando	30	24	72	6.5	3.53	10.8	79.5	3694	13.1	33.8	12.4	11	4	11	61	941.181
Tamba	31	45	85	7.3	3.22	11.4	80.7	1844	11.5	48.1	18.5	1	1	1	53	891.708
Atlanta	31	24	72	9	3.37	10.9	82.8	3226	5.1	45.2	12.3	5	3	10	61	871.338
Augusta	42	40	77	6.1	3.45	10.4	71.8	2269	22.7	41.4	19.5	8	3	5	53	971.122
Columbus	43	27	72	9	3.25	11.5	87.1	2909	7.2	51.6	9.5	7	3	10	56	887.466
Macon	46	55	84	5.6	3.35	11.4	79.7	2647	21	46.9	17.9	6	5	1	59	952.529
Savannah	39	29	75	8.7	3.23	11.4	78.6	4412	15.6	46.6	13.2	13	7	33	60	968.665
Chicago	35	31	81	9.2	3.1	12	78.3	3262	12.6	48.6	13.9	7	4	4	55	919.729
Rockford	43	32	74	10.1	3.38	9.5	79.2	3214	2.9	43.7	12	11	7	32	54	844.053
Gary	11	53	68	9.2	2.99	12.1	90.6	4700	7.8	48.9	12.3	648	319	130	47	861.833
Indianapolis	30	35	71	8.3	3.37	9.9	77.4	4474	13.1	42.6	17.7	38	37	193	57	989.265
South Bend	50	42	82	7.3	3.49	10.4	72.5	3497	36.7	43.3	26.4	15	18	34	59	1006.49
Terre Haute	60	67	82	10	2.98	11.5	88.6	4657	13.5	47.3	22.4	3	1	1	60	861.439
Des Moines	30	20	69	8.8	3.26	11.1	85.4	2934	5.8	44	9.4	33	23	125	64	929.15
Topeka	25	12	73	9.2	3.28	12.1	83.1	2095	2	51.9	9.8	20	11	26	58	857.622
Wichita	45	40	80	8.3	3.32	10.1	70.3	2682	21	46.1	24.1	17	14	78	56	961.009
Baton Rouge	46	30	72	10.2	3.16	11.3	83.2	3327	8.8	45.3	12.2	4	3	8	58	923.234
New Orleans	54	54	81	7.4	3.36	9.7	72.8	3172	31.4	45.5	24.2	20	17	1	62	1113.156
Shreveport	42	33	77	9.7	3.03	10.7	83.5	7462	11.3	48.7	12.4	41	26	108	58	994.648
Portland	42	32	76	9.1	3.32	10.5	87.5	6092	17.5	45.3	13.2	29	32	161	54	1015.023
Baltimore	36	29	72	9.5	3.32	10.6	77.6	3437	8.1	45.5	13.8	45	59	263	56	991.29
Boston	37	38	67	11.3	2.99	12	81.5	3387	3.6	50.3	13.5	56	21	44	73	893.991
Brockton	42	29	72	10.7	3.19	10.1	79.5	3508	2.2	38.8	15.7	6	4	18	56	938.5
Fall River	41	33	77	11.2	3.08	9.6	79.9	4843	2.7	38.6	14.1	11	11	89	54	946.185
Springfield	44	39	78	8.2	3.32	11	79.9	3768	28.6	49.5	17.5	12	9	48	53	1025.502
Worcester	32	25	72	10.9	3.21	11.1	82.5	4355	5	46.4	10.8	7	4	18	60	874.281
Detroit	34	32	79	9.3	3.23	9.7	76.8	5160	17.2	45.1	15.3	31	15	68	57	953.56
Flint	10	55	70	7.3	3.11	12.1	88.9	3033	5.9	51	14	144	66	20	61	839.709
Jackson	18	48	63	9.2	2.92	12.2	87.7	4253	13.7	51.2	12	311	171	86	71	911.701
Lansing	13	49	68	7	3.36	12.2	90.7	2702	3	51.9	9.7	105	32	3	71	790.733
Saginaw	35	40	64	9.6	3.02	12.2	82.5	3626	5.7	54.3	10.1	20	7	20	72	899.264
Duluth	45	28	74	10.6	3.21	11.1	82.6	1883	3.4	41.9	12.3	5	4	20	56	904.155
Minneapolis	38	24	72	9.8	3.34	11.4	78	4923	3.8	50.5	11.1	8	5	25	61	950.672
Jackson	31	26	73	9.3	3.22	10.7	81.3	3249	9.5	43.9	13.6	11	7	25	59	972.464
Kansas City	40	23	71	11.3	3.28	10.3	73.8	1671	2.5	47.4	13.5	5	2	11	60	912.202
St. Louis	41	37	78	6.2	3.25	12.3	89.5	5308	25.9	59.7	10.3	65	28	102	52	967.803
Omaha	28	32	81	7	3.27	12.1	81	3665	7.5	51.6	13.2	4	2	1	54	823.764
Las Vegas	45	33	76	7.7	3.39	11.3	82.2	3152	12.1	47.3	10.9	14	11	42	56	1003.502
Manchester	45	24	70	11.8	3.25	11.1	79.8	3678	1	44.8	14	7	3	8	56	895.696
Atlantic City	42	33	76	9.7	3.22	9	76.2	9699	4.8	42.2	14.5	8	8	49	54	911.817

Appendix II: Final Dataset

Cities	MORT	ln(SO2)	JANT	EDUC	NONW	
Birmingham	921.87	4.0775374	27	11.4	8.8	
Mobile	997.875	3.6635616	23	11	3.5	
Montgomery	962.354	3.4965076	29	9.8	0.8	
Phoenix	982.291	3.1780538	45	11.1	27.1	
Little Rock	1071.289	5.3278762	35	9.6	24.4	
Fresno	1030.38	4.2766661	45	10.2	38.5	
Los Angeles	934.7	4.1271344	30	12.1	3.5	
Sacramento	899.529	1.3862944	30	10.6	5.3	
San Diego	1001.902	3.6109179	24	10.5	8.1	
San Francisco	912.347	2.9957323	27	10.7	6.7	
San Jose	1017.613	3.2958369	42	9.6	22.2	
Denver	1024.885	5.6276211	26	10.9	16.3	
Bridgeport	970.467	4.9836066	34	10.2	13	
Hartford	985.95	4.1588831	28	11.1	14.7	
New Haven	958.839	2.7080502	31	11.9	13.1	
Wilmington	860.101	0	46	11.8	14.8	
Washington	936.234	2.7725887	30	11.4	12.4	
Jacksonville	871.766	3.3322045	30	12.2	4.7	
Miami	959.221	4.8202816	27	10.8	15.8	
Orlando	941.181	2.3978953	24	10.8	13.1	
Tamba	891.708	0	45	11.4	11.5	
Atlanta	871.338	2.3025851	24	10.9	5.1	
Augusta	971.122	1.6094379	40	10.4	22.7	
Columbus	887.466	2.3025851	27	11.5	7.2	
Macon	952.529	0	55	11.4	21	
Savannah	968.665	3.4965076	29	11.4	15.6	
Chicago	919.729	1.3862944	31	12	12.6	
Rockford	844.053	3.4657359	32	9.5	2.9	
Gary	861.833	4.8675345	53	12.1	7.8	
Indianapolis	989.265	5.2626902	35	9.9	13.1	
South Bend	1006.49	3.5263605	42	10.4	36.7	
Terre Haute	861.439	0	67	11.5	13.5	
Des Moines	929.15	4.8283137	20	11.1	5.8	
Topeka	857.622	3.2580965	12	12.1	2	
Wichita	961.009	4.3567088	40	10.1	21	
Baton Rouge	923.234	2.0794415	30	11.3	8.8	
New Orleans	1113.156	0	54	9.7	31.4	
Shreveport	994.648	4.6821312	33	10.7	11.3	
Portland	1015.023	5.0814044	32	10.5	17.5	
Baltimore	991.29	5.572154	29	10.6	8.1	
Boston	893.991	3.7841896	38	12	3.6	
Brockton	938.5	2.8903718	29	10.1	2.2	
Fall River	946.185	4.4886364	33	9.6	2.7	

Springfield	1025.502	3.871201011	39	11	28.6	
Worcester	874.281	2.890371758	25	11.1	5	
Detroit	953.56	4.219507705	32	9.7	17.2	
Flint	839.709	2.995732274	55	12.1	5.9	
Jackson	911.701	4.454347296	48	12.2	13.7	
Lansing	790.733	1.098612289	49	12.2	3	
Saginaw	899.264	2.995732274	40	12.2	5.7	
Duluth	904.155	2.995732274	28	11.1	3.4	
Minneapolis	950.672	3.218875825	24	11.4	3.8	
Jackson	972.464	3.218875825	26	10.7	9.5	
Kansas City	912.202	2.397895273	23	10.3	2.5	
St. Louis	967.803	4.624972813	37	12.3	25.9	
Omaha	823.764	0	32	12.1	7.5	
Las Vegas	1003.502	3.737669618	33	11.3	12.1	
Manchester	895.696	2.079441542	24	11.1	1	
Atlantic City	911.817	3.891820298	33	9	4.8	
Jersey City	954.442	3.663561646	28	10.7	11.7	

Appendix III: Descriptive Analysis and Five Number Summary of Each Attribute

1) DENS

Five-Number Summary	
Minimum	1441
First Quartile	3042
Median	3567
Third Quartile	4657
Maximum	9699

Descriptive Summary

	DENS
Mean	3876.05
Median	3567
Mode	#N/A
Minimum	1441
Maximum	9699
Range	8258
Variance	2114413.6754
Standard Deviation	1454.1024
Coeff. of Variation	37.52%
Skewness	1.3795
Kurtosis	3.5910
Count	60
Standard Error	187.7238

2) EDUC

Five-Number Summary	
Minimum	9
First Quartile	10.4
Median	11.05
Third Quartile	11.5
Maximum	12.3

	EDUC
Mean	10.97333333
Median	11.05
Mode	11.4
Minimum	9
Maximum	12.3
Range	3.3
Variance	0.7145
Standard Deviation	0.8453
Coeff. of Variation	7.70%
Skewness	-0.2249
Kurtosis	-0.7513
Count	60
Standard Error	0.1091

3) HC

Five-Number Summary	
Minimum	1
First Quartile	7
Median	14.5
Third Quartile	31
Maximum	648

	HC
Mean	37.85
Median	14.5
Mode	6
Minimum	1
Maximum	648
Range	647
Variance	8459.8924
Standard Deviation	91.9777
Coeff. of Variation	243.01%
Skewness	5.5934
Kurtosis	34.6852
Count	60
Standard Error	11.8743

4) HOUS

Five-Number Summary	
Minimum	66.8
First Quartile	78.3
Median	81.15
Third Quartile	83.9
Maximum	90.7

	HOUS
Mean	80.91333333
Median	81.15
Mode	79.9
Minimum	66.8
Maximum	90.7
Range	23.9
Variance	26.4337
Standard Deviation	5.1414
Coeff. of Variation	6.35%
Skewness	-0.4170
Kurtosis	0.3849
Count	60
Standard Error	0.6637

5) HUMID

Five-Number Summary	
Minimum	38
First Quartile	55
Median	57
Third Quartile	60
Maximum	73

	HUMID
Mean	57.66666667
Median	57
Mode	56
Minimum	38
Maximum	73
Range	35
Variance	28.8362
Standard Deviation	5.3699
Coeff. of Variation	9.31%
Skewness	0.2375
Kurtosis	4.2895
Count	60
Standard Error	0.6933

6) JANT

Five-Number Summary	
Minimum	12
First Quartile	27
Median	31.5
Third Quartile	40
Maximum	67

	JANT
Mean	33.98333333
Median	31.5
Mode	30
Minimum	12
Maximum	67
Range	55
Variance	103.4065
Standard Deviation	10.1689
Coeff. of Variation	29.92%
Skewness	0.9607
Kurtosis	1.0878
Count	60
Standard Error	1.3128

7) JULT

Five-Number Summary	
Minimum	63
First Quartile	72
Median	74
Third Quartile	78
Maximum	85

JULT

Mean	74.58333333
Median	74
Mode	72
Minimum	63
Maximum	85
Range	22
Variance	22.6879
Standard Deviation	4.7632
Coeff. of Variation	6.39%
Skewness	0.1367
Kurtosis	0.0109
Count	60
Standard Error	0.6149

8) Ln(SO2)

Five-Number Summary	
Minimum	0
First Quartile	2.397895
Median	3.39897
Third Quartile	4.276666
Maximum	5.627621

	ln(SO@)
Mean	3.197212975
Median	3.398970206
Mode	0
Minimum	0
Maximum	5.627621114
Range	5.627621114
Variance	2.2428
Standard Deviation	1.4976
Coeff. of Variation	46.84%
Skewness	-0.7145
Kurtosis	0.0164
Count	60
Standard Error	0.1933

9) NONW

Five-Number Summary	
Minimum	0.8
First Quartile	4.8
Median	10.4
Third Quartile	15.8
Maximum	38.5

	NONW
Mean	11.87
Median	10.4
Mode	13.1
Minimum	0.8
Maximum	38.5
Range	37.7
Variance	79.5869
Standard Deviation	8.9211
Coeff. of Variation	75.16%
Skewness	1.1311
Kurtosis	0.9360
Count	60
Standard Error	1.1517

10) NOX

Five-Number Summary	
Minimum	1
First Quartile	4
Median	9
Third Quartile	26
Maximum	319

	NOX
Mean	22.65
Median	9
Mode	4
Minimum	1
Maximum	319
Range	318
Variance	2146.7737

Standard Deviation	46.3333
Coeff. of Variation	204.56%
Skewness	5.1656
Kurtosis	30.2637
Count	60
Standard Error	5.9816

11) OVR65

Five-Number Summary	
Minimum	5.6
First Quartile	7.6
Median	9
Third Quartile	9.7
Maximum	11.8

	OVR65
Mean	8.798333333
Median	9
Mode	9.2
Minimum	5.6
Maximum	11.8
Range	6.2
Variance	2.1449
Standard Deviation	1.4646
Coeff. of Variation	16.65%
Skewness	-0.0341
Kurtosis	-0.6037
Count	60
Standard Error	0.1891

12) POOR

Five-Number Summary	
Minimum	9.4
First Quartile	12
Median	13.2
Third Quartile	15.3
Maximum	26.4

	POOR
Mean	14.37333333
Median	13.2
Mode	13.2
Minimum	9.4
Maximum	26.4
Range	17
Variance	17.3064
Standard Deviation	4.1601
Coeff. of Variation	28.94%
Skewness	1.4635
Kurtosis	1.5314
Count	60
Standard Error	0.5371

13) POPN

Five-Number Summary	
Minimum	2.92
First Quartile	3.21
Median	3.265
Third Quartile	3.36
Maximum	3.53

	POPN
Mean	3.263166667
Median	3.265
Mode	3.21
Minimum	2.92
Maximum	3.53
Range	0.61
Variance	0.0183
Standard Deviation	0.1353
Coeff. of Variation	4.14%
Skewness	-0.4893
Kurtosis	0.0438
Count	60
Standard Error	0.0175

14) PREC

Five-Number Summary	
Minimum	10
First Quartile	32
Median	38
Third Quartile	44
Maximum	60

	PREC
Mean	37.36666667
Median	38
Mode	36
Minimum	10
Maximum	60
Range	50
Variance	99.6938
Standard Deviation	9.9847
Coeff. of Variation	26.72%
Skewness	-0.8022
Kurtosis	1.2836
Count	60
Standard Error	1.2890

15) WWDRK

Five-Number Summary	
Minimum	33.8
First Quartile	43.1
Median	45.5
Third Quartile	49.6
Maximum	59.7

	WWDRK
Mean	46.08166667
Median	45.5
Mode	42.6
Minimum	33.8
Maximum	59.7
Range	25.9
Variance	21.2802
Standard Deviation	4.6130
Coeff. of Variation	10.01%
Skewness	0.0985
Kurtosis	0.5549
Count	60
Standard Error	0.5955

16) MORT

Five-Number Summary	
Minimum	790.733
First Quartile	895.696
Median	943.683
Third Quartile	985.95
Maximum	1113.156

	MORT
Mean	940.3584333
Median	943.683
Mode	#N/A
Minimum	790.733
Maximum	1113.156
Range	322.423
Variance	3869.6211
Standard Deviation	62.2063
Coeff. of Variation	6.62%
Skewness	0.0984
Kurtosis	0.1633
Count	60
Standard Error	8.0308